DOI:
10.1016/j.jaci.2014.06.032

Document Version
Peer reviewed version

IL-9 is a key component of memory Th cell peanut-specific responses from peanut allergic children – author post-print final draft (posted under Creative Commons Non-Commercial No Derivatives License)

Helen A. Brough, MSc FRCPCH,[1,2] David J. Cousins, PhD,[1,2,3] Alina Munteanu, MSc,[4] Yuen Fei Wong, PhD,[5] Asha Sudra, MSc,[1,2] Kerry Makinson, MSc,[1,2] Alick C. Stephens, PhD,[1,2] Matthew Arno, PhD,[5] Liviu Ciortuz, PhD,[4] Gideon Lack, FRCPCH MD,[1,2] Victor Turcanu,[1],2 MD PhD

[1]Division of Asthma, Allergy and Lung Biology, [2]MRC & Asthma UK Centre in Allergic Mechanisms of Asthma, King's College London, Guys' Hospital, London, SE1 9RT

[2] Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, LE3 9QP, UK.

[3]Faculty of Computer Science, University of Iasi, Romania.

[4]Genomics Centre, King's College London UK,

Correspondence:

Dr Victor Turcanu MD PhD

Division of Asthma, Allergy and Lung Biology

King's College London

5th Floor Bermondsey Wing

Guys' Hospital SE1 9RT UK

Telephone: +44 (0)2071888716

Fax: +44 (0)2074038690

Email: victor.turcanu@kcl.ac.uk

Funding:

This work was supported by the National Peanut Board, USA and by the Medical Research Council UK, Asthma UK, and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Abbreviations:

B7: $\alpha_4\beta_7$ integrin

CISH: Cytokine-inducible SH2-containing protein

CLA: Cutaneous lymphocyte antigen

FACS: Flow-cytometry

ICCS: Intracellular cytokine staining

IQR: Interquartile range

NA: non-peanut allergic

OVA: ovalbumin

PA: peanut allergic

PBMC: Peripheral blood mononuclear cell

PBS: Phosphate buffered saline

PS: Peanut sensitized

RF: Random Forrest algorithm

RT-qPCR: Real time quantitative polymerase chain reaction

TGFβ: Transforming Growth Factor beta

Th: T helper

Word count: 3595

Abstract:

Background: Differentiation between peanut allergic (PA) and peanut sensitized (PS) individuals who tolerate peanut but have peanut-specific IgE and/or positive skin prick tests represents a significant diagnostic difficulty. Previously, gene expression microarrays were successfully used to identify biomarkers and explore immune responses during peanut allergy immunotherapy.

Objective: Characterization of peanut-specific responses from PA, PS and atopic non-peanut-allergic (NA) children.

Methods: A preliminary, exploratory microarray investigation of gene expression in peanut-activated memory Th subsets from 3PA and 3NA children identified potential peanut allergy diagnostic biomarkers. Microarray findings were confirmed using RT-qPCR in 30 individuals (12PA, 12PS and 6NA). Flow cytometry was used to identify the Th subsets involved.

Results: Amongst 387 differentially expressed genes, IL9 showed the greatest difference between PA and NA (45.59-fold change, $p<0.001$) followed by IL5 then IL13. Notably, IL9 allowed the most accurate classification of PA and NA using a machine learning approach using recursive feature elimination and the Random Forest algorithm. Skin- and respectively gut-homing Th cells from PA donors expressed similar Th2 and Th9-associated genes. RT-qPCR confirmed that IL9 was the highest differentially expressed gene between PA and NA (23.3-fold change, $P<0.01$) and PS children (18.5-fold change, $P<0.05$). Intracellular cytokine staining showed that IL9 and the Th2-specific cytokine IL5 are produced by distinct Th populations.

Conclusion: In this study, IL9 best differentiated between PA and PS (and atopic NA) children. Mutually exclusive production of IL9 and the Th2-specific IL5 suggests that the IL9-producing cells belong to the recently described Th9 subset.

Clinical applicability: IL9 may be a useful biomarker to distinguish between peanut allergic and peanut sensitized children. Further prospective studies are required to validate these findings and determine whether IL9 responses could be targeted in treating and preventing peanut allergy.

Capsule summary: In vitro peanut-specific responses of memory Th cells from peanut allergic children include IL9 production in addition to the well-characterized Th2 cytokines.

Key words: Peanut allergy, peanut sensitization, gene expression, RT-qPCR, microarray, IL9, Th9, IL5, IL13, Th2

Abstract word count: 246

Introduction

Peanut allergy has increased over the last decades so that its prevalence in childhood is currently estimated at 1.4%[1] in the USA and reaches 1.8% in the UK.[2;3] Since peanut allergy is rarely outgrown and is responsible for a significant proportion of severe anaphylactic reactions to foods, it represents a major population health concern.[4]

Mechanistically, peanut allergy is driven by Th cells whose determinant role was confirmed by the observation that allergy can be transferred from peanut allergic (PA) donors to non-peanut allergic (NA) recipients via solid organ transplantation which entails Th cell transfer.[5] Conversely, peanut allergy resolution has been observed after bone marrow transplantation.[6] We[7] and others[8;9] showed that the in vitro peanut-specific Th cell response in PA individuals is dominated by a Th2-polarised population characterized by the production of Th2 cytokines such as IL4, IL5 and IL13. However, we also observed that a significant number of in vitro peanut-responding Th cells did not produce Th2 (nor Th1-specific) cytokines even when restimulated with phorbol ester and calcium ionophore, suggesting that they may belong to distinct Th cell subsets.[10]

Indeed, the discovery of Treg, Th17, Th22 and more recently Th9 shows that Th cell responses go beyond the binary Th1/Th2 paradigm.[11] Treg cells suppress IgE production[12] and block mast cell, basophil and eosinophil activation,[13] whereas in mouse models of asthma, Th17 cells increase eosinophilia and IgE production.[14] Th9 was recently identified as a distinct Th cell subset induced by the combination of IL4 and TGFβ.[15] IL9, which is the Th9 subset-defining cytokine, is a mast cell growth factor which increases cytokine production in activated mast cells and enhances IL4-driven IgE production by B cells. IL9 is highly expressed in the lungs of asthmatic humans[16] and in an experimental mouse model, neutralizing anti-IL9 antibodies were shown to ameliorate Th9-mediated asthma.[17;18]

In the present study, we aimed to further dissect the peanut-responding Th cell subsets by separating the skin-homing Th cells which express the Cutaneous Lymphocyte Antigen (CLA)[19] and the gut-homing memory Th cells (which express the B7 gut-homing marker) respectively.[20] Skin exposure to peanut antigens following skin stripping to mimic an eczema phenotype leads to strong Th2-skewed responses in mice, supporting the concept of epicutaneous sensitization.[21] Furthermore, we recently showed that peanut-specific proliferation is quantitatively higher in the skin-homing than in gut-homing Th cells in PA individuals.[22] Conversely, oral exposure to antigens underlies oral tolerance and is associated with higher levels of antigen-specific proliferation of the gut-homing T cells.[20] This has led to the Dual Allergen Exposure Hypothesis being proposed, in that the timing and balance of cutaneous versus oral exposure to an allergen may determine whether the child develops allergy or tolerance.[23]

In this study we aimed to assess differential gene expression in peripheral blood mononuclear cells (PBMCs) cultured with peanut from PA, peanut sensitized (PS) and atopic NA children, in both skin and gut homing Th cells and flow-cytometry (FACS) analysis of intracellular cytokine production

Materials and methods

Study participants

PA, PS and atopic NA donors were recruited from a tertiary referral allergy center at St. Thomas Hospital Children's Allergy Unit, London UK. Informed consent was obtained prior to participation. Ethical approval for this study was obtained from St Thomas Research Ethics Committee (ref.10/H0802/45). Three PA and three atopic NA children were recruited for gene expression analysis and 36 microarrays were carried out on various T cell populations derived from their peanut-stimulated and unstimulated PBMCs (as negative controls). Participants recruited for the gene microarray work were all males in order to avoid differential gene expression linked to gender. Forty patients recruited for subsequent reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) and FACS analysis did not have gender as a selection criterion. Peanut allergy was defined as having a convincing history of an immediate hypersensitivity reaction following exposure to peanut and a skin prick test to peanut extract of ≥8mm or peanut-specific (s)IgE ≥15kU/L.[24] Peanut tolerance was defined as regular peanut consumption, i.e. eating an age-appropriate amount of peanut in one meal without any allergic symptoms within the last month; of which some were PS (defined by SPT≥3mm or peanut (s)IgE ≥0.35kU/L). If there was no history of peanut consumption, children underwent a double-blind placebo controlled food challenge (DBPCFC). The demographic and clinical characteristics of all study participants are described in Table E1.

Details of the methodology are described in the online repository, with an overview below:

1) PBMC isolation and peanut stimulation in vitro
2) Cell labeling and flow-cytometry sorting
3) RNA extraction
4) cDNA synthesis
5) cDNA fragmentation, biotin-labeling and hybridization

6) Microarray data normalization and quality checks

7) Partek Suite analysis of microarray

8) Microarray gene expression classification approach

9) RT-qPCR confirmation of gene expression

10) FACS analysis of intracellular cytokine production

Overview of methods

PBMCs from 3 PA and 3 NA children were cultured with peanut antigens for 18h. Subsequently activated (CD69+) skin-homing (CLA+) memory Th cells (CD4+CD45RO), and respectively activated (CD69+) gut-homing (B7+) memory Th cells were sorted by FACS into Trizol LS. mRNA was extracted using the miRNeasy Mini kit (Qiagen, UK) and cDNA was synthesized using the NuGEN Ovation® Pico WTA system v2 kit (NuGEN, California, USA). Following data normalization and quality checks, differential gene expression between PA and NA donors was determined by Affymetrix GeneChip® Human Gene 1.0 ST Array. Microarray results were analyzed using Partek Genomics Suite and using an automatic classification approach, as currently used in data mining / machine learning fields. Microarray findings were subsequently confirmed at a gene level using RT-qPCR on whole PBMCs from 30 children (12 PA, 12 PS and 6 NA) and at a protein level using FACS intracellular cytokine staining (ICCS) in 5 PA and 5 PS individuals.

Results

Microarray analysis of gene expression in peanut-activated memory Th cells from PA vs. NA donors.

The level of gene expression in peanut-activated memory Th cells from PA and atopic NA donors is displayed as a volcano plot using Partek Genomics Suite (*FIG.* 1); the horizontal axis shows the fold-change difference of gene expression between PA and NA whereas the vertical axis displays the statistical significance of the difference of gene expression between PA and atopic NA, taking into account gene expression variability within the PA and the NA groups respectively. The genes appearing in the top right area were upregulated in PA versus NA; the most statistically significant differently expressed gene was IL9 (45.59 fold change, $p<0.001$) followed by the Th2 gene cluster containing IL5 and IL13. Conversely, the genes appearing in the top left area (e.g. SULT1B1, RGS18) were upregulated in atopic NA donors.

Peanut specific memory T-helper cells from PA donors express typical Th2-specific genes as well as Th9 associated genes

A hierarchical clustering heatmap of the dataset (*FIG.* 2) consisting of 48 differentially expressed genes, selected through a filter criteria of at least two-fold changes of expression with $P<0.05$ was produced using Partek Genomics Suite. There was notable differential expression between peanut stimulated activated (CD69+) Th cells from PA and NA donors for Th2 signature genes (IL-4, IL-5, IL-13) and Th9 associated genes (IL-9, IRF4, IL17RB).[25] CCL1 (Chemokine (C-C motif) ligand 1), IL31 and IL3 were preferentially expressed in skin-homing Th cells from PA donors. Other genes differentially expressed in PA versus NA and their functions and involvement in allergy are described in Table I.[25-40]Most genes preferentially expressed in NA vs. PA were as yet not functionally understood and only SULT1B1, GCNT4 and RGS18 could be identified (Table I).

ANOVA of *genes expressed in activated skin and gut-homing peanut-responding memory Th cells from PA donors vs. non-activated PA Th cells, non-antigen PA Th cells and NA Th cells*

In order to compare the basal, non-stimulated level of gene expression in skin and gut homing peanut-activated memory Th cells, we performed 36 microarrays (displayed in Table E2) in three types of Th cell subsets isolated from PA and NA donors respectively:

I) Peanut activated: CD69+ Th cells from PBMCs cultured in peanut protein

II) Peanut non-activated: CD69- Th cells that were not activated isolated from the same PBMCs cultured with peanut protein (internal negative controls)

III) No peanut (unstimulated): CD69- Th cells isolated from the PBMCs where no peanut antigen was added (external negative controls)

Differential gene expression of peanut stimulated (P+) activated (CD69+) Th cells between PA and atopic NA is displayed in *FIG.* 3; we also include results for internal negative controls (peanut stimulated, non-activated (CD69-)) and external negative controls (unstimulated, non-activated (CD69-)) in the online repository (*FIG.* E4). There was differential gene expression of IL9 and associated genes IL17RB and IRF4 as well as classical Th2 cytokines (IL5, IL13 and IL4) and CISH (Cytokine-inducible SH2-containing protein) in PA versus NA donors. CCL1 and IL31 were preferentially increased in skin-homing CD69+ memory Th2 cells from PA children; however, Th2/Th9 cytokines were not expressed preferentially in skin- or gut-homing Th cells. All external negative controls (unstimulated CD69- Th cells) in PA and NA donor had low expression of all genes assessed. There was very little background expression of IL9, IL4, IL5, IL13, IL31 and CCL1 in internal negative controls (P+ CD69-) in PA and NA donors; however, there was some background expression of IL17RB, IRF4 and CISH in P+ CD69- Th cells (*FIG.* E4).

IL9 allows the differentiation between PA and NA individuals when used in the frame of a Random Forest (RF) classifier.

We used RF to classify the peanut stimulated, activated (CD69+) Th cell responses in PA vs. NA children. We built 250, 500 and 750 trees (analysis pathways); the number of analysis pathways is dictated by a trade-off between the efficiency (computing speed and memory consumption) of the classifier and the quality of its predictions. Using this approach RF was able to differentiate between PA vs. NA children with 100% accuracy even when the gene selection was limited to sets of 10, then 5, then only 2 genes. In this final analysis, the two genes that are able to accurately distinguish between PA and NA samples were: IL9 and HSPA5 for 250 trees, IL9 and 8069610 for 500 trees, and IL9 and HUWE1 for 750 trees (*FIG.* E5). From these sets of genes IL9 expression was the best discriminating gene between memory Th cells from PA and NA children..

Differential gene expression of Th2 and Th9 cytokines between PA and NA individuals is confirmed by RT-qPCR

The expression of IL5 and IL13 (signature genes for Th2 responses) and respectively IL9 and IRF4 (reflecting Th9 responses) was higher in PBMCs isolated from PA donors than NA donors (FIG. 4). Relative quantitation (RQ) for IL9 was 25.40 [Interquartile range [IQR] 4.09-65.33] in PA versus 1.09 [IQR 0.96-2,10] in NA; which equated to a 23.30-fold change in IL9 expression. RQ for IL5 was 61.73 [IQR 4.29-167.87] in PA versus 2.86 [1.80-4.97] in NA (21.6-fold change); RQ for IL13 was 53.96 [IQR 31.68-121.14] in PA versus 4.54 [IQR 1.80-7.15] in NA (11.89-fold difference). Thus IL9 had the highest fold differential gene expression between PA and NA donors. Significant differential gene expression (RQ) was also observed for other genes known to be involved in allergic responses such as IL31, CISH and CCL1 ($P<0.01$) and the Th2/Th9-specific transcription factor IRF4 ($P<0.05$). On RT-qPCR analysis there was no significant differential IL4 and IL17RB expression between PA and NA donors.

Differential gene expression of Th2 and Th9 cytokines between PA and PS individuals by RT-qPCR

Following the finding that IL9 had the highest differential expression in PA versus NA we sought to determine whether this would also be a useful biomarker to differentiate between PA and PS children (*FIG.* 4). RQ for IL9 was 25.40 [IQR 4.09-65.33] in PA versus 1.37 [IQR 1.21-3.88] in PS (18.54-fold change). RQ for IL5 was 61.73 [IQR 4.29-167.87] in PA versus 3.40 [1.14-14.29] in PS (18.16-fold change) and for IL13 was 53.96 [IQR 31.68-121.14] in PA versus 7.56 [IQR 4.08-22.86] in PS (7.14-fold difference). Thus IL9 also had the highest fold differential gene expression between PA and PS donors. CCL1, CISH and IL31 were also differentially expressed between PA and PS donors. One PS child who was eating whole peanuts 1-2 times/month had higher IL9, IL5 and IL31 gene expression levels which represents the green outlier points for these three RQ box-plots (Fig 5).

Different Th populations secrete IL9 and IL5 cytokines

ICCS of peanut stimulated memory T-helper cells demonstrated that IL5 and IL9 are produced by distinct Th-cell subsets (FIG. 5A and 6B and Table E3); memory Th cells from all PA donors produced more IL9 and IL5 than PS children and the pattern of cytokine expression was mutually exclusive.

Discussion

We employed the well-established Affymetrix microarray method to determine gene expression profiles of memory T cells responding to peanut antigens in vitro in order to characterize differential gene expression and the different Th cell subsets involved in peanut allergy. Microarray gene expression data found evidence of high IL9 expression in activated memory Th cells from PA donors in addition to the expected Th2 gene signature. Applying a Random Forrest algorithm we found that IL9 was the best gene classifier that allows differentiation between PA and NA children with 100% accuracy. Using RT-qPCR and FACS we confirmed the presence of IL9 at a gene and protein level, and showed that IL9 was the best gene to distinguish between PA and NA children (23.3-fold difference) as well as between PA and PS children (18.5-fold difference); thus, making IL9 a good biomarker for clinical peanut allergy in this study. The differential expression of Th2 signature genes (IL5 and IL13) and Th9 associated genes (IL9 and the transcription factor IRF4) in PA versus NA allergic children as well as the dichotomous pattern of IL5 and IL9 production in peanut-stimulated activated memory T cells suggests the involvement of both Th2 and a distinct Th9 subset in peanut allergy.

Jabeen et al. reviewed the recently discovered Th cell subset Th9 that produces IL9 but not the other Th2-defining cytokines such as IL5 and IL13.[41] Th9 cells have been described in the inflamed skin of patients with atopic eczema[42] and in the bronchial mucosa in mouse models of allergic asthma,[16] and Th9 is emerging as an important T cell subset in human respiratory allergic disease.[15;41] IL9 has previously been described as the top ranking gene (from 1482 differentially expressed genes) for discriminating between atopic and non-atopic responses to house dust mite.[37] IL9 production is also important for in vivo allergic responses in seasonal allergic rhinitis, as successful specific immunotherapy with grass pollen led to the decrease of IL9 mRNA and IL9 protein in the patients' nasal mucosa.[43] IRF4 is necessary for the differentiation of Th9 cells,[29;30] and has previously been described

317 using gene expression microarrays in allergen-stimulated PBMCs from patients with allergic
318 rhinitis.[44]
319
320 In mice, oral antigen induced anaphylaxis to ovalbumin (OVA) is IgE mediated and
321 predominantly IL9 and IL9 receptor pathway dependent.[45] $IL9^{-/-}$ and $IL9R^{-/-}$ mice developed
322 OVA-specific IgE levels following intraperitoneal OVA/alum immunisation but did not develop
323 anaphylaxis after OVA oral gavage and had reduced intestinal mast cell proliferation and
324 degranulation. This might explain the role of IL9 in IgE mediated food allergy versus only IgE
325 specific sensitization. In a recent study in humans, IL9 was the best gene to discriminate
326 between peanut stimulated PBMCs from PA versus NA (28-fold difference on RT-qPCR);
327 however, IL9 expression was not assessed in PS individuals.[46] Other differences between
328 their manuscript and ours were (i) they assessed adults rather than children, (ii) their peanut
329 tolerant group were generally non-atopic whereas the majority of our peanut tolerant group
330 were atopic thus accounting for genes upregulated due to atopy rather than peanut allergy
331 perse (iii) exploratory microarray analysis was not used to determine candidate genes thus
332 other important differentiation genes might have been excluded (iv) IL9 gene expression was
333 initially investigated in the entire PBMC population rather than in different T cells subsets to
334 determine which cells were producing IL-9; however, importantly, they showed significantly
335 raised IL5 and IL9 levels in the 5-day peanut culture supernatants of PA versus NA
336 individuals (using ELISA).
337
338 We confirmed differential gene expression between PA and NA for IL9 and its transcription
339 factor IRF4, signature Th2 genes (IL5, IL13) and other genes important for allergen specific
340 responses (CISH, IL31 and CCL1) by RT-qPCR. Confirmation of microarray findings was
341 performed using RT-qPCR of whole PBMCs rather than FACS sorted cells because, if IL9 is
342 to be useful as a diagnostic biomarker then it should be able to distinguish between PA and
343 NA individuals in unseparated cells rather than in very small Th cell subsets that involve
344 cumbersome and highly skilled experimental procedures. The IL25 receptor (IL17RB) and

IL4 were no longer significantly differentially expressed between PA and NA on RT-qPCR. The IL25 receptor (IL17RB) is expressed in Th9 cells generated in vitro in the presence of TGFβ and IL4.[47] The difference between findings in the microarray and RT-qPCR may be because the microarray was performed on peanut cultured activated memory Th cells, whereas the RT-qPCR was performed on whole PBMCs thus the signal may not be as strong in a mixed cell culture; furthermore, on microarray analysis IL17RB and IL4 were not as highly differentially expressed as IL9, IL5 and IL13 in PA versus NA (*FIG.* 1). CISH was upregulated in PA versus NA and PS children and has been shown to be differentially expressed in house dust mite-stimulated T cells from atopic individuals vs. non-atopic individuals[37;48] and in ovalbumin-stimulated PBMCs from egg allergic vs. non-egg allergic patients.[40]

A notable finding in this study was the lack of differential expression of Th2 and Th9-related genes between skin and gut-homing memory Th cell subsets in PA individuals given our previous findings.[22] Additionally, there was no significant difference in the number CLA+ versus B7+ activated memory Th cells sorted from peanut stimulated culture in PA versus NA children, although there was a trend towards higher responses to peanut from the gut-homing versus skin-homing memory Th subset in NA donors (P=0.05). One potential explanation for this is that the PA individuals whom we investigated had long-term, well established peanut allergy; it is therefore plausible that whilst the original Th2 and Th9 gene expression profile may have been limited only to the skin-homing Th cell subset, in the long term the Th2 and Th9 responses became dominant in all lymphocyte homing compartments as a consequence of allergic individuals' repeated exposure to peanut, either through eczematous skin or through accidental oral exposure. Future experiments could further identify the differences between peanut-specific immune responses of skin and gut-homing subsets in PS individuals, if this was carried out longitudinally, in a prospective approach, as children progress from PS to PA.

373 Nonetheless, other genes were differentially expressed in skin vs. gut homing Th cells;
374 CCL1 (Chemokine C-C motif ligand 1), IL31 and IL3 were upregulated in skin homing Th
375 cells from PA donors compared with gut-homing Th cells from PA and gut- and skin-homing
376 Th cells from NA donors. CCL1 is secreted by activated T cells and IgE-activated human
377 mast cells and binds to the CCR8 receptor expressed by Th2 cells, dendritic cells,
378 monocytes, NK and immature B cells.[41] Notably, CCR8 is expressed by approximately 70%
379 of Th cells recruited in the asthmatic airways, which is why oral forms of the small molecule
380 CCR8 antagonists are currently being developed for therapeutic purposes.[39] IL31 is
381 produced mainly by activated Th2 cells and is increased in skin biopsies of patients with
382 eczema and contact dermatitis.[25] IL3 increases the activation and release of mediators from
383 eosinophils and basophils in response to IgE FCεRI cross-linking. The differential expression
384 of these genes reflects the contrast between the skin and gut homing Th cell subsets
385 observed in the PCA analysis (*FIG*. E3); however, the differential expression of homing-
386 related genes was not mirrored by a differential expression of Th2 and Th9 subset-defining
387 cytokine genes. The lack of such difference suggests that in patients with well-established
388 peanut allergy, allergen-specific T cell populations comprise Th2 and Th9 cells regardless of
389 their homing phenotype.
390
391 Another unexpected finding was the absence of an identifiable Treg gene signature in the
392 peanut-responding memory Th cells from NA donors; however, the microarray did identify
393 three genes SULT1B1, GCNT4 and RGS18 and other as yet unidentified genes which were
394 differentially expressed in NA vs. PA children. The absence of classical Treg biomarkers
395 suggests that a suppressor cytokine environment is not actively induced following only short-
396 term (18h) in vitro stimulation with antigens from tolerated foods. Another potential
397 explanation for this finding is that peanut specific (CD69+) Treg cells in NA do not express
398 CLA or B7 homing markers and were thus not gated and isolated on FACS analysis.
399

Conclusion

In this study, IL9 was the best discriminator for PA children versus NA children and also importantly PS children. Mutually exclusive production of Th9 specific (IL9) and Th2 specific (IL5) cytokines suggest that IL9-producing cells belong to the distinct Th9 subset population. The use of gene expression microarrays to generate hypotheses by evaluating the overall immune response to an allergen in a small number of patients, followed by further in-depth investigations has previously been successfully applied to elucidate the mechanisms underlying peanut oral immunotherapy,[49] and for predicting the efficacy of venom immunotherapy.[50] Future research into the interplay between the Th9 and Th2 subsets may clarify whether the success of preventative therapeutic approaches aimed at peanut allergy resolution could be evaluated on the basis of IL9 secretion and/or Th9 suppression in peanut-specific in vitro responses. Prospective studies should further evaluate IL9 as a biomarker for PA and as a potential target for the prevention and treatment of PA.

Acknowledgements: The authors are grateful to Professor Patrick G Holt from the Telethon Institute for Child Research, Perth, Australia for many helpful discussions and advice when designing this project. We thank Bogdan Luca from the University of Iasi for his help running the data mining classifiers on the microarray data. We are also grateful to Dr Susanne Heck, PJ Chana and Helen Graves from the NIHR flow cytometry core laboratory at Guy's Hospital, London UK for their support with FACS sorting and analysis.

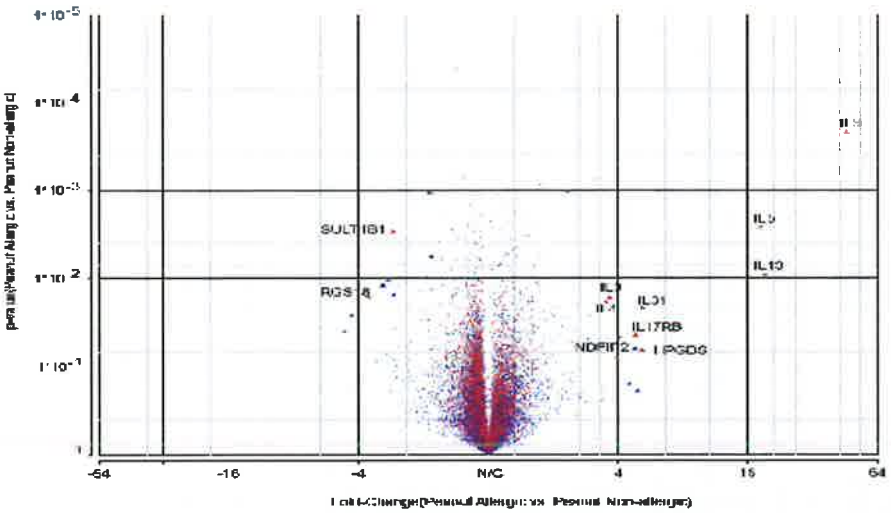
Table I. List and outline of the known roles in allergic responses of genes which were significantly different ($p<0.05$) between PA and NA donors. Genes are listed in order of statistical significance. All genes were upregulated in PA vs. NA except the SULT1B1, GCNT4 and RGS18 which were upregulated in NA vs. PA (grey fill).

Gene	p-value	Biological functions of genes and involvement in allergy
IL9	0.0002	IL9 is produced by Th9 cells, mast cells and eosinophils; its main targets are T cells, B cells and mast cells. TGFβ in the presence of IL4 reprograms Th2-cell differentiation and leads to the development of Th9 cells that produce IL-9 and IL-10. Thus IL-4 blocks the generation of TGFβ–induced Foxp3+ Treg cells and induces Th9 cells. IL-9 is involved in Th2 inflammatory reactions, promotes the production of IL4–induced IgE, induces chemokine and mucus secretion by bronchial epithelial cells, and leads to mast cell proliferation. [25]
IL5	0.0024	IL5 is an eosinophil and B-cell growth factor mainly produced by Th2 cells, activated eosinophils, mast cells, CD8+Tc2 cells, NK cells and NKT cells. IL5 promotes proliferation, activation, differentiation, survival and adhesion of eosinophils. [25]
SULT1B1	0.0029	Sulfotransferase family cytosolic 1B member 1 enzyme catalyzes the sulfate conjugation of hormones, neurotransmitters and drugs.[26]
IL13	0.0086	IL13 is secreted by activated Th2 cells, mast cells, basophils, eosinophils and NKT cells. IL13 targets are B cells, mast cells, epithelial cells, eosinophils, smooth muscle cells and macrophages. IL-13 contributes to asthma and allergic rhinitis late-phase responses.[25]
IL4	0.0134	IL4 is produced mainly byTh2 cells, basophils, eosinophils, mast cells and a subset of NK T cells. The main IL4 targets are T and B cells to induce the differentiation of antigen-stimulated naive T cells into Th2 cells, increase the expression of class II MHC molecules in B cells, enhance expression of CD23, induce IgE class switch and upregulate expression of IL4R. [25]
LIMA1	0.0135	LIM domain and actin binding 1 is a cytoskeleton-associated protein that inhibits actin filament depolymerization.[27]
IL31	0.0137	IL31 is a cytokine related to the IL6–type cytokines. IL31 is produced mainly by activated Th2 cells and CD8+ T cells. IL31 expression is increased in skin biopsies of patients with eczema and allergic contact dermatitis. Serum IL31 levels is correlated with eczema severity.[25]
IL3	0.0149	IL3 is expressed by T cells, macrophages, stromal cells, NK cells, mast cells, and eosinophils. Amongst other functions IL3 also increases the activation and release of mediators from eosinophils and basophils in response to IgE FcεR cross-linking.[25]
CD200R1	0.0153	Cell surface glycoprotein CD200 receptor 1 is a receptor for the OX-2 membrane glycoprotein. Chronic infection drives expression of the inhibitory receptor CD200R, and its ligand CD200, by mouse and human CD4 T cells.[36]

RGS18	0.0176	Regulator of G-protein signaling 18 is a protein implicated in lymphocyte function by modulating signaling through 7-transmembrane receptors, in chemokine signaling and homing, and in differentiation and transformation.[28]
FAM115A	0.0237	Family with sequence similarity 115, member A is a protein with unknown function. [33]
IRF4	0.0243	Interferon regulatory factor 4 regulates a large number of genes relevant to Th cell differentiation.[29] IRF4 is also essential for the developmental program of Th9 cells.[30]
CEACAM1	0.0277	Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) is a cell-cell adhesion molecule detected on leukocytes, epithelia and endothelia. Multiple cellular activities including apoptosis, tumor suppression and modulation of innate and adaptive immune responses.[31]
IL4R	0.0281	There are 2 types of IL4 binding receptors: type I IL4R, which predominates in hematopoietic cells and which is responsible for IL4 signaling in T cells, and type II IL4R, which is expressed on hematopoietic and non-hematopoietic cells but not T cells.[25]
GCNT4	0.0289	Glucosaminyl (N-acetyl) transferase 4 is an enzyme which mediates core 2 O-glycan branching, important in mucin-type biosynthesis.[32]
MAL	0.0300	Myelin and lymphocyte protein has been localized to the endoplasmic reticulum of T-cells and is a candidate linker protein in T-cell signal transduction. It is increased in house-dust mite-stimulated Th cells from atopic individuals.[37]
CCL1	0.0337	Chemokine (C-C motif) ligand 1 is secreted by activated T cells and IgE-activated human and mast cells and binds to the CCR8 receptor. CCR8 is expressed by dendritic cells, monocytes, NK and immature B cells and is associated with Th2 cells.[38]
CISH	0.0385	Cytokine-inducible SH2-containing protein controls interleukin-2 signaling; and can be induced by IL2, IL3, GM-CSF and EPO in hematopoietic cells. CISH is increased in HDM-stimulated T cells from atopic individuals,[37] and ovalbumin-stimulated PBMC from egg allergic patients. [40]
IL17RB	0.0410	IL17RB binds IL17B and IL17E (IL25). IL17 amplifies delayed-type reactions by increasing chemokine production to recruit monocytes and neutrophils to the site of inflammation.[34]
MS4A3	0.0486	Membrane-spanning 4-domains subfamily A member 3 likely plays a role in signal transduction in lymphoid cells. MS4A3 is increased in patients with latex and/or vegetable food allergy.[35]

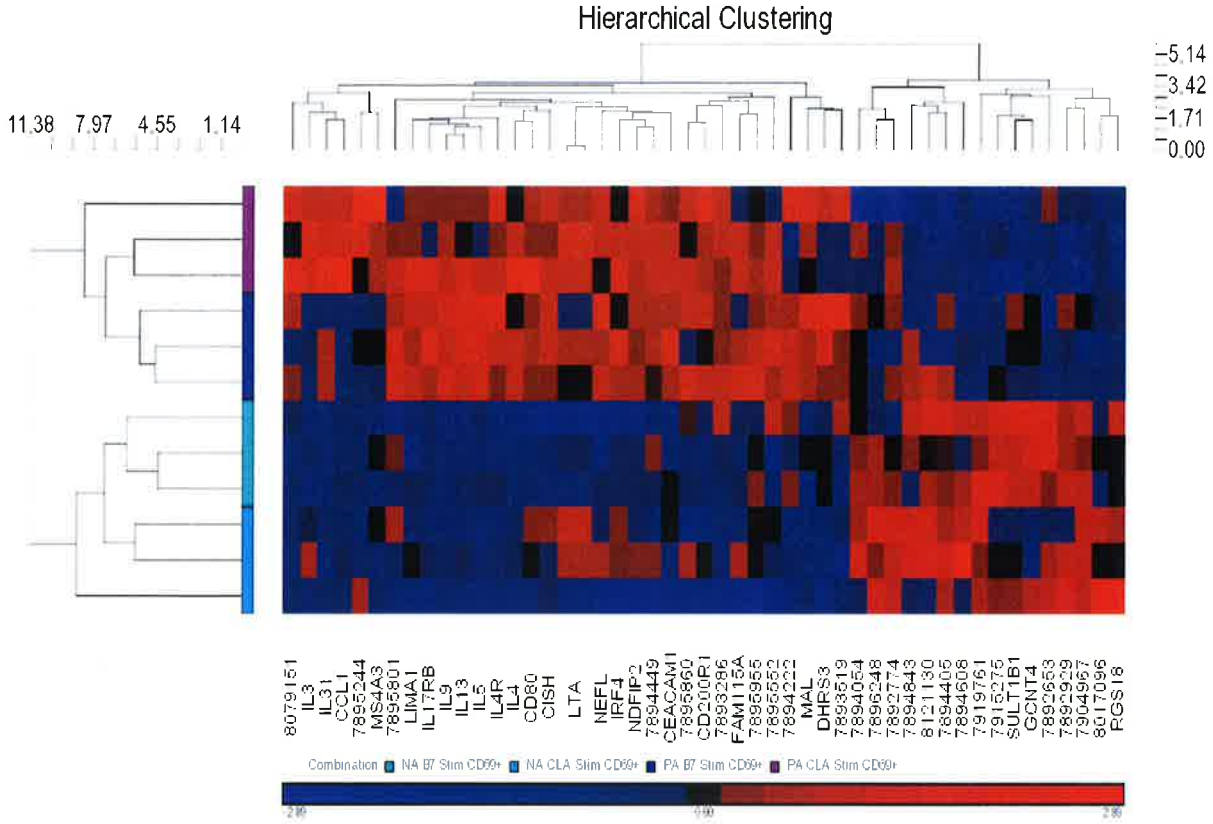
423

424 **Figure legends:**



425

426 FIG. 1. Volcano plot of gene expression fold-change of significantly differentially expressed
427 genes in peanut-activated memory Th cells from 3 PA and 3 NA donors. Horizontal axis:
428 Fold-change differential gene expression between PA and NA (positive values indicate
429 higher expression in PA, negative values reflect higher expression in NA). Vertical axis:
430 Statistical significance of differential gene expression between PA and NA, adjusting for
431 gene expression variability within PA and NA groups respectively. Points represent individual
432 genes.



433

434 FIG. 2. Heatmap of 48 differentially-expressed genes in skin and gut-homing peanut-
435 activated memory Th cells from PA (n=3) and NA donors (n=3) (≥2 fold changes of
436 expression, $p<0.05$). Each column represents Th cell gene expression from an individual
437 donor and each row shows a single gene. Upregulated genes are red ('hot') whereas
438 downregulated genes are blue ('cold').

439

440 FIG. 3. Expression of Th2/Th9 subset-specific genes and other genes known to be involved
441 in allergic responses in skin (CLA+) and gut-homing (B7+) peanut-activated memory Th cells
442 from PA (n=3) and NA (n=3) donors.

443 FIG 4: Differential RT-qPCR gene expression in PBMCs from PA (n=12), PS (n=12) and
444 atopic NA(n=6) children comparing target gene relative to the endogenous gene 18s (ΔCT)
445 in peanut stimulated versus unstimulated cultures (ΔΔCT) converted into relative quantitation
446 (RQ) in log transformed arbitrary units (AU).

447

448 FIG. 5. Box-plot of FACS sorted memory Th cells and intracellular cytokine staining of IL5
449 and IL9 expressed in unstimulated versus peanut-stimulated memory Th cells from 5 PA and
450 5 PS donors (A) and an example of FACS analysis of peanut stimulated memory Th cells
451 from the first PA and PS donor (B).

452
453

References

(1) Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. J Allergy Clin Immunol 2010; 125(6):1322-6.

(2) Du Toit G, Katz Y, Sasieni P, Mesher D, Maleki SJ, Fisher HR et al. Early consumption of peanuts in infancy is associated with a low prevalence of peanut allergy. J Allergy Clin Immunol 2008; 122(5):984-91.

(3) King RM, Knibb RC, Hourihane JO. Impact of peanut allergy on quality of life, stress and anxiety in the family. Allergy 2009; 64(3):461-8.

(4) Bock SA, Munoz-Furlong A, Sampson HA. Further fatalities caused by anaphylactic reactions to food, 2001-2006. J Allergy Clin Immunol 2007; 119(4):1016-8.

(5) Schuller A, Barnig C, Matau C, Geny S, Gosselin M, Moal MC et al. Transfer of peanut allergy following lung transplantation: a case report. Transplantation Proceedings 2011; 43(10):4032-5.

(6) Hourihane JO, Rhodes HL, Jones AM, Veys P, Connett GJ. Resolution of peanut allergy following bone marrow transplantation for primary immunodeficiency. Allergy 2005; 60(4):536-7.

(7) Turcanu V, Winterbotham M, Kelleher P, Lack G. Peanut-specific B and T cell responses are correlated in peanut-allergic but not in non-allergic individuals. Clin Exp Allergy 2008;(7):1132-9.

(8) Flinterman AE, Pasmans SG, Den Hartog Jager CF, Hoekstra MO, Bruijnzeel-Koomen CA, Knol EF et al. T cell responses to major peanut allergens in children with and without peanut allergy. Clin Exp Allergy 2010; 40(4):590-7.

(9) Higgins JA, Lamb JR, Lake RA, O'Hehir RE. Polyclonal and clonal analysis of human CD4+ T-lymphocyte responses to nut extracts. Immunology 1995; 84(1):91-7.

(10) Turcanu V, Maleki SJ, Lack G. Characterization of lymphocyte responses to peanuts in normal children, peanut-allergic children, and allergic children who acquired tolerance to peanuts. J Clin Invest 2003; 111(7):1065-72.

(11) Jutel M, Akdis CA. T-cell subset regulation in atopy. [Review]. Current Allergy & Asthma Reports 2011; 11(2):139-45.

(12) Mittag D, Scholzen A, Varese N, Baxter L, Paukovics G, Harrison LC et al. The effector T cell response to ryegrass pollen is counterregulated by simultaneous induction of regulatory T cells. J Immunol 2010; 184(9):4708-16.

(13) Palomares O, Yaman G, Azkur AK, Akkoc T, Akdis M, Akdis CA. Role of Treg in immune regulation of allergic diseases. [Review] [82 refs]. Eur J Immunol 2010; 40(5):1232-40.

(14) Wakashin H, Hirose K, Maezawa Y, Kagami S, Suto A, Watanabe N et al. IL-23 and Th17 cells enhance Th2-cell-mediated eosinophilic airway inflammation in mice. Am J Resp Crit Care Med 2008; 178(10):1023-32.

493 (15) Stassen M, Schmitt E, Bopp T. From interleukin-9 to T helper 9 cells.
494 [Review]. Annals of the New York Academy of Sciences 2012; 1247:56-68.

495 (16) Erpenbeck VJ, Hohlfeld JM, Volkmann B, Hagenberg A, Geldmacher H,
496 Braun A et al. Segmental allergen challenge in patients with atopic asthma leads to
497 increased IL-9 expression in bronchoalveolar lavage fluid lymphocytes. J Allergy Clin
498 Immunol 2003; 111(6):1319-27.

499 (17) Kung TT, Luo B, Crawley Y, Garlisi CG, Devito K, Minnicozzi M et al. Effect of
500 anti-mIL-9 antibody on the development of pulmonary inflammation and airway
501 hyperresponsiveness in allergic mice. Am J Respir Cell Mol Biol 2001; 25(5):600-5.

502 (18) Cheng G, Arima M, Honda K, Hirata H, Eda F, Yoshida N et al. Anti-
503 interleukin-9 antibody treatment inhibits airway inflammation and hyperreactivity in mouse
504 asthma model. Am J Resp Crit Care Med 2002; 166(3):409-16.

505 (19) Santamaria Babi LF, Picker LJ, Perez Soler MT, Drzimalla K, Flohr P, Blaser
506 K et al. Circulating allergen-reactive T cells from patients with atopic dermatitis and allergic
507 contact dermatitis express the skin-selective homing receptor, the cutaneous lymphocyte-
508 associated antigen. J Exp Med 1995; 181(5):1935-40.

509 (20) Kantele A, Zivny J, Hakkinen M, Elson CO, Mestecky J. Differential homing
510 commitments of antigen-specific T cells after oral or parenteral immunization in humans. J
511 Immunol 1999; 162(9):5173-7.

512 (21) Strid J, Hourihane J, Kimber I, Callard R, Strobel S. Epicutaneous exposure
513 to peanut protein prevents oral tolerance and enhances allergic sensitization. Clin Exp
514 Allergy 2005; 35(6):757-66.

515 (22) Chan SM, Turcanu V, Stephens AC, Fox AT, Grieve AP, Lack G. Cutaneous
516 lymphocyte antigen and alpha4beta7 T-lymphocyte responses are associated with peanut
517 allergy and tolerance in children. Allergy 2012; 67(3):336-42.

518 (23) Lack G. Epidemiologic risks for food allergy. J Allergy Clin Immunol
519 2008;(6):1331-6.

520 (24) Roberts G, Lack G. Diagnosing peanut allergy with skin prick and specific IgE
521 testing. J Allergy Clin Immunol 2005;(6):1291-6.

522 (25) Akdis M, Burgler S, Crameri R, Eiwegger T, Fujita H, Gomez E et al.
523 Interleukins, from 1 to 37, and interferon-gamma: receptors, functions, and roles in diseases.
524 J Allergy Clin Immunol 2011; 127(3):701-21.

525 (26) Meinl W, Glatt H. Structure and localization of the human SULT1B1 gene:
526 neighborhood to SULT1E1 and a SULT1D pseudogene. Biochemical & Biophysical
527 Research Communications 2001; 288(4):855-62.

528 (27) Xu X, Guo J, Vorster P, Wu Y. Involvement of LIM kinase 1 in actin
529 polarization in human CD4 T cells. Communicative & Integrative Biology 2012; 5(4):381-3.

530 (28) Gagnon AW, Murray DL, Leadley RJ. Cloning and characterization of a novel
531 regulator of G protein signalling in human platelets. Cellular Signalling 2002; 14(7):595-606.

(29) Staudt V, Bothur E, Klein M, Lingnau K, Reuter S, Grebe N et al. Interferon-regulatory factor 4 is essential for the developmental program of T helper 9 cells. Immunity 2010; 33(2):192-202.

(30) Ahyi AN, Chang HC, Dent AL, Nutt SL, Kaplan MH. IFN regulatory factor 4 regulates the expression of a subset of Th2 cytokines. J Immunol 2009; 183(3):1598-606.

(31) Nagaishi T, Iijima H, Nakajima A, Chen D, Blumberg RS. Role of CEACAM1 as a regulator of T cells. [Review] [83 refs]. Annals of the New York Academy of Sciences 2006; 1072:155-75.

(32) Furukawa K, Sato T. Beta-1,4-galactosylation of N-glycans is a complex process. [Review] [59 refs]. Biochimica et Biophysica Acta 1999; 1473(1):54-66.

(33) Swat A, Dolado I, Igea A, Gomez-Lopez G, Pisano DG, Cuadrado A et al. Expression and functional validation of new p38 transcriptional targets in tumorigenesis. Biochemical Journal 2011; 434(3):549-58.

(34) Parmentier CN, Fuerst E, McDonald J, Bowen H, Lee TH, Pease JE et al. Human T(H)2 cells respond to cysteinyl leukotrienes through selective expression of cysteinyl leukotriene receptor 1. J Allergy Clin Immunol 2012; 129(4):1136-42.

(35) Saulnier N, Nucera E, Altomonte G, Rizzi A, Pecora V, Aruanno A et al. Gene expression profiling of patients with latex and/or vegetable food allergy. European Review for Medical & Pharmacological Sciences 2012; 16(9):1197-210.

(36) Caserta S, Nausch N, Sawtell A, Drummond R, Barr T, Macdonald AS et al. Chronic infection drives expression of the inhibitory receptor CD200R, and its ligand CD200, by mouse and human CD4 T cells. PLoS ONE 2012; 7(4):e35466.

(37) Bosco A, McKenna KL, Devitt CJ, Firth MJ, Sly PD, Holt PG. Identification of novel Th2-associated genes in T memory responses to allergens. J Immunol 2006; 176(8):4766-77.

(38) Gonzalo JA, Qiu Y, Lora JM, Al-Garawi A, Villeval JL, Boyce JA et al. Coordinated involvement of mast cells and T cells in allergic mucosal inflammation: critical role of the CC chemokine ligand 1:CCR8 axis. J Immunol 2007; 179(3):1740-50.

(39) Connolly S, Skrinjar M, Rosendahl A. Orally bioavailable allosteric CCR8 antagonists inhibit dendritic cell, T cell and eosinophil migration. Biochemical Pharmacology 2012; 83(6):778-87.

(40) Nakajima Y, Tsuge I, Kondo Y, Komatsubara R, Hirata N, Kakami M et al. Up-regulated cytokine-inducible SH2-containing protein expression in allergen-stimulated T cells from hen's egg-allergic patients. Clin Exp Allergy 2008; 38(9):1499-506.

(41) Jabeen R, Kaplan MH. The symphony of the ninth: the development and function of Th9 cells. [Review]. Current Opinion in Immunology 2012; 24(3):303-7.

(42) Cortelazzi C, Campanini N, Ricci R, De PG. Inflammed skin harbours Th9 cells. Acta Dermato-Venereologica 2013; 93(2):183-5.

(43) Nouri-Aria KT, Pilette C, Jacobson MR, Watanabe H, Durham SR. IL-9 and c-Kit+ mast cells in allergic rhinitis during seasonal allergen exposure: effect of immunotherapy. J Allergy Clin Immunol 2005; 116(1):73-9.

(44) Bruhn S, Barrenas F, Mobini R, Andersson BA, Chavali S, Egan BS et al. Increased expression of IRF4 and ETS1 in CD4+ cells from patients with intermittent allergic rhinitis. Allergy 2012; 67(1):33-40.

(45) Osterfeld H, Ahrens R, Strait R, Finkelman FD, Renauld JC, Hogan SP. Differential roles for the IL-9/IL-9 receptor alpha-chain pathway in systemic and oral antigen-induced anaphylaxis. J Allergy Clin Immunol 2010; 125(2):469-76.

(46) Xie J, Lotoski LC, Chooniedass R, Su RC, Simons FE, Liem J et al. Elevated antigen-driven IL-9 responses are prominent in peanut allergic humans. PLoS ONE [Electronic Resource] 2012; 7(10):e45377.

(47) Angkasekwinai P, Chang SH, Thapa M, Watarai H, Dong C. Regulation of IL-9 expression by IL-25 signaling. Nature Immunology 2010; 11(3):250-6.

(48) Bosco A, McKenna KL, Firth MJ, Sly PD, Holt PG. A network modeling approach to analysis of the Th2 memory responses underlying human atopic disease. J Immunol 2009; 182(10):6011-21.

(49) Jones SM, Pons L, Roberts JL, Scurlock AM, Perry TT, Kulis M et al. Clinical efficacy and immune regulation with peanut oral immunotherapy. J Allergy Clin Immunol 2009; 124(2):292-300.

(50) Niedoszytko M, Bruinenberg M, de MJ, Wijmenga C, Platteel M, Jassem E et al. Gene expression analysis in predicting the effectiveness of insect venom immunotherapy. J Allergy Clin Immunol 2010; 125(5):1092-7.

Online Repository

PBMC isolation and peanut stimulation in vitro

Anticoagulated blood was centrifuged for 5 minutes at 620g at room temperature in order to collect autologous plasma. After the plasma layer was collected, cells were diluted with a volume of PBS equal to the volume of plasma removed, and then decanted into cell separation tubes over a Histopaque-1077 layer. These were centrifuged at 1200g for 15 minutes at room temperature then the PBMC at the interface were collected and washed three times with PBS. The PBMC were counted using a Nucleocounter, then were resuspended in serum free AIM V medium (containing penicillin and streptomycin) + 2-Mercaptoethanol (both from Invitrogen Life Technologies, UK). PBMC were cultured for 18 hours in order to detect the early allergen-specific response as previously described.[E1] Cells were distributed in 24-well plates at a concentration of 3 million cells in 1ml per well in order to ensure optimal cell-to-cell contact [E2] and were either unstimulated or stimulated with 200 micrograms/ml whole defatted peanut protein extract (ALK Abello, Horsholm, Denmark).

Cell labeling and flow-cytometry sorting

Cells were aspirated from the peanut-stimulated and unstimulated cultures respectively, and were washed and stained on ice for 20 minutes with antibodies specific for CD4 as Th cell marker (CD4-APC,BD Biosciences, Oxford UK) and CD45RO as memory cell marker (CD45RO Pacific Blue, Biolegend, London UK). CLA which is an adhesion molecule expressed by skin homing cells was used as skin homing marker (CLA-FITC, Biolegend). Beta 7 integrin which is expressed by gut homing cells was used as agut homing marker (B7 PerCP/Cy5.5, Biolegend). CD69 labeling (CD69-PECy7, Biolegend) was used as a marker of cell activation to identify peanut-responding cells because its expression is rapidly upregulated on T cells post activation and remains stable for up to 72h.[E3] Thus, activated and non-activated, skin- and gut-homing memory Th cells (i.e. respectively CD4+CD45RO+CLA+B7-CD69+, CD4+CD45RO+CLA+B7-CD69-, CD4+CD45RO+CLA+

B7-CD69+ and CD4+CD45RO+CLA-B7+CD69-) were 4-way sorted from both peanut-stimulated and unstimulated PBMC cultures using a FACSAria™ II Cell Sorter. We display the FACS sorting experiment for all 3 PA and 3 NA donors in *FIG.* E1. Selected cells were sorted directly into TRIzol LS Reagent (Life Technologies Corp., UK) with a 3:1 ratio of TRIzol LS Reagent to volume of sample. Samples were vortexed and stored at -80°C until RNA extraction was performed. The number of cells FACS sorted for each condition are displayed in Table E2.

RNA extraction

RNA was extracted using the miRNeasy Mini kit (Qiagen, UK) according to the manufacturer's protocol, using TRIzol LS instead of QIAzol. We performed a DNase digestion step to remove any traces of contaminating DNA and carried out the water elution of the RNA from the column twice to enhance the RNA yield. The amount and quality of RNA and cDNA obtained from the FACS sorted cells from PA and NA donors are displayed in Table E2.

cDNA synthesis

The quality and quantity of total RNA were first examined on an Agilent 2100 Bioanalyser using the Agilent RNA 6000 Pico Kit (Agilent Technologies, Wokingham, UK). Upon confirming that the RNA integrity was of good quality, cDNA was synthesized using the NuGEN Ovation® Pico WTA system v2 kit (NuGEN, San Carlos, California, USA) following the manufacturer's protocol. Briefly, 500 pg - 50 ng of total RNA in 5 μl water was reverse transcribed into first-strand cDNA. Second-strand cDNA was then synthesized, which served as a template for amplification of transcripts through a proprietary Ribo-SPIA® amplification step.

cDNA fragmentation, biotin-labeling and hybridization

3 μg of amplified cDNA was fragmented and biotin-labeled using the NuGEN Encore™ Biotin Module following the manufacturer's protocol. The hybridization cocktail was then prepared from the fragmented, biotin-labeled cDNA using the Affymetrix Hybridization, Wash and Stain kit and hybridized to the Affymetrix GeneChip® Human Gene 1.0 ST Array for 18 hours at 45 ° C and 60rpm rotation. Upon completion of hybridization, washing and staining steps were performed on an Affymetrix GeneChip® Fluidics Station 450 and arrays were scanned using an Affymetrix GeneChip® 3000 7G scanner.

Microarray data normalization and quality checks

A Robust Multichip Average (RMA) method was used for data normalization. The array signal before and after normalization is shown in *FIG.* E2 Ai and Aii. All arrays passed the basic quality checks, including a uniform hybridizing signal of the spike-in hybridization control at the staggered concentration of CreX>BioD>BioC>BioB (FIG E2 B), and a high positive vs. negative area under the curve ("pos_vs_neg_auc") that ranged between 0.82 and 0.90 (data not shown). Primary component analysis (PCA) of peanut-induced gene expression was carried out using the Partek Genomics Suite software (St Louis, Missouri, USA) on the normalized data from the peanut-activated memory Th cells; n=6 (3 CLA+ and 3 B7+ samples) microarrays in PA and n=6 (3 CLA+ and 3 B7+ samples) microarrays in NA. PCA shows the variation in all the genes expressed in one 3 dimensional image. The Partek software took all genes (variables) and plotted them all against each other, then ranked genes into orders of variability (genes which do not change very much are ranked low and genes which change are ranked high). Genes were then assembled into groups of genes with similar variability of variation. Groups of genes with the most variability were grouped together (PC1) followed by PC2 then PC3. Thus the PCA resulted in three arms of groups of genes based on the variability of gene expression and the similarity of patterns of gene expression. CLA+ and B7+ samples from PA donors (red) and NA donors (blue) are shown in *FIG.* E3. The first three components (PC1=20.7%, PC2=14.3% and PC3=10.7%) account for 45.7% of global differential gene expression. Applying PCA on these samples leads to a

clear-cut clustering of CLA+ skin-homing Th cells (green centroid), distinguishing them from the B7 gut-homing Th cells (yellow centroid) on the basis of their gene expression.

Partek Suite analysis

In order to create a Heat-map of differentially expressed genes we used a 4-way ANOVA using peanut allergic status, Th homing (skin versus gut), donor (to adjust for donor effect) and date of microarray (to adjust for batch effect). The Heatmap (*FIG.* 1) was based on all genes with 2-fold differential expression and *P*-value <0.05. The actual *P*-values obtained using this 4-way ANOVA are described in Table 1 and dot-plots for individual differential gene expression are shown in *FIG.* 3. In order perform a detailed analysis of differential gene expression in the 36 PA and 36 NA samples (including internal and external negative controls), we subsequently used a 6-way ANOVA using stimulation (peanut stimulated versus unstimulated), CD69+ activation status as well as peanut allergic status, Th homing (skin versus gut), donor and date of microarray. The dot plots for individual gene expression using the 6-way ANOVA analysis are displayed in the supplementary online repository (*FIG.* E2).

Microarray gene expression classification approach (FIG. E5)

Microarray data was analyzed by an automatic classification approach, as currently used in data mining / machine learning fields; in these types of analyses the number of samples to be worked upon is usually much larger than the number of data points ("features") used to describe those samples. Since in our study the number of genes (features) in each microarray (r=32,020) was much larger than the number of microarrays (n=36) (samples) a basic filter was first applied, based on the fact that most genes show very small variability in their expression levels as they are involved in the basal cellular metabolism. The inter-quartile range (IQR) of individual gene expression in all six patients (three PA and three NA) was used to identify those genes whose expression levels vary most between the two allergy phenotypes. The global inter-quartile range (IQR) of gene expression in each microarray

was also calculated and any genes with a variability that was smaller than 20% of that value were removed. This filter was applied using the open-source statistical software R (□www.r-project.org/□), and it narrowed down the analysis from 32,020 to 12,257 genes, representing about one third of all genes.

Using an ANOVA filter for statistical comparison, the analysis was further narrowed down to a 387 subset of genes differentially expressed between the peanut-stimulated skin-homing and gut homing samples from PA and NA donors respectively. Using this set of 387 differentially expressed genes, Random Forest (RF), [E4] which is a state-of-the-art supervised classification algorithm, together with recursive feature elimination, was used to classify all the peanut-activated samples into PA vs. NA groups. The Random Forrest classifier is an ensemble of multiple classification trees (analysis pathways) and is suitable for our analysis with only six samples since its performance is estimated directly, without the need for cross-validation, by the out-of-bag error fraction. The accuracy of each set of genes compared between samples was computed by subtracting this error from 1, its maximum value. Recursive feature elimination is an algorithm used to select a subset of relevant data points or "features" (e.g. gene expression levels) for use in model construction; it is usually used with support vector machines to repeatedly remove features with low weights when constructing a classification model. [E5]

An advantage of RF is the fact that it can measure each gene's ("feature's") "importance" relative to the classification task at hand. Although there are several kinds of such importance measures, the unscaled permutation importance was used in this case, because it is recommended for feature selection. [E6] It measures, for each gene, the decrease of classification accuracy in the case of random permutation of its expression levels, averaged over all trees. The feature selection process that we used in combination with the algorithm is itself a recursive meta-algorithm in which, at each step, a random forest of decision trees was grown and then 50% of the genes (i.e. those with the smallest differential gene

expression level) were discarded. For a given sample, at each step of the classification process, a feature (i.e. gene) was analyzed and, based on its value; a decision was reached to assign that sample to a certain class. The gene selection stopped when the number of remaining genes in the set used for classification fell below a certain, initially fixed threshold. All classification tasks were performed using the free software package Random Jungle. [E7]

Real-time quantitative polymerase chain reaction (RT-qPCR)

In order to confirm the microarray findings, RNA was extracted from peanut-stimulated PBMC obtained from an independent group of 13 PA and 17 peanut tolerant donors (of which 11 were peanut sensitized). Briefly, PBMC were cultured for 18 hours serum free AIM V medium (as described above) in the presence of 200 micrograms/ml whole defatted peanut protein extract (ALK Abello) or without any peanut antigen added (as an unstimulated negative control). RT-qPCRs were performed with the following TaqMan MGB probes (Applied Biosystems, Invitrogen, Paisley, United Kingdom), as previously described: [E8] IL3, Hs00174117_m1; IL4, Hs00929862_m1; IL5, Hs00174200_m1; IL9, Hs00174125_m1; IL13, Hs00174379_m1; IL31, Hs01098710_m1; CCL1, Hs00171072_m1; CISH, Hs00367082_g1; IRF4, Hs01056533_m1; 18s rRNA, 4319413F. Target gene expression levels were compared against endogenous control 18s ribosomal RNA (ΔCT) then against the peanut unstimulated control ($\Delta\Delta$CT) and converted to RQ ($2^{-(\Delta\Delta CT)}$). Genes with no amplification were assigned a CT value of 40. The Mann-Whitney test was performed RQ values between NA and PA and between PS and PA using SPSS (SPSS 19.0; SPSS Inc., Chicago, IL, USA).

FACS analysis of intracellular cytokine production

Following results obtained from the gene microarray and RT-qPCR, we sought to confirm our gene expression findings on a protein expression level. PBMC from 5 PA and 5 PS donors were cultured overnight in RPMI 1640 supplemented with 5% autologous plasma in the presence of 200 micrograms/ml whole defatted peanut protein extract (ALK Abello) or without any antigen added (as an unstimulated negative control) in the presence of

monensin (2 micrograms/ml) and brefeldin A (1 microgram/ml) as inhibitors of cytokine secretion. After incubation, PBMC were fixed, permeabilized, stained with antibodies specific for CD4, CD45RO, CLA, B7, IL5 and IL9 and cytokine production (IL5 and IL9) was analyzed by FACS. IL-5 was used to identify Th2 cells whereas IL-9 was used as the hallmark cytokine to identify the Th9 subset.

References

(E1) Bosco A, McKenna KL, Devitt CJ, Firth MJ, Sly PD, Holt PG. Identification of novel Th2-associated genes in T memory responses to allergens. J Immunol 2006; 176(8):4766-77.

(E2) Turcanu V, Maleki SJ, Lack G. Characterization of lymphocyte responses to peanuts in normal children, peanut-allergic children, and allergic children who acquired tolerance to peanuts. J Clin Invest 2003; 111(7):1065-72.

(E3) Alari-Pahissa E, Notario L, Lorente E, Vega-Ramos J, Justel A, Lopez D et al. CD69 does not affect the extent of T cell priming. PLoS ONE [Electronic Resource] 2012; 7(10):e48593.

(E4) Breiman L. Random Forests. Machine Learning. 2001: 5-32.

(E5) Zhang X, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN et al. Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinformatics 2006; 7:197.

(E6) Diaz-Uriarte R, Alvarez de AS. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006; 7:3.

(E7) Schwarz DF, Konig IR, Ziegler A. On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data.[Erratum appears in Bioinformatics. 2011 Feb 1;27(3):439]. Bioinformatics 2010; 26(14):1752-8.

(E8) Parmentier CN, Fuerst E, McDonald J, Bowen H, Lee TH, Pease JE et al. Human T(H)2 cells respond to cysteinyl leukotrienes through selective expression of cysteinyl leukotriene receptor 1. J Allergy Clin Immunol 2012; 129(4):1136-42.

196 Table E1. Clinical characteristics of PA and NA donors recruited for gene microarray analysis (n=6), FACS intracellular cytokine staining (n=7
197 and RT-qPCR (n=30). The term N/A is stated where this information is not available.

	Peanut allergic status	Sex	Age (yrs)	Ethnicity	Last known peanut exposure	Allergic reaction	Skin prick test (mm)	Peanut specific IgE (kU/L)	Ara h 2	Ara h 1	Ara h 8	Current eczema	Eczema severity	Asthma	Atopic disease
Gene expression analysis using Affimetrix microarray	PA	M	9	Caucasian	7 years	Hives, facial angioedema, cough, difficulty breathing	10	>100	>100	>100	0.1	Yes	Mild	Yes	Allergic rhinitis (AR), tree nut allergy
	PA	M	7	Black-British	4 years	Hives, angiodema	12	31.8	213.3	38	27	Yes	Moderate	Yes	Fish allergy, AR, asthma
	PA	M	15	Caucasian	14 years	Generalised hives, cough and vomit	22	>100	68	40	N/A	No	Severe	Yes	AR, tree nut sensitised
	NA	M	8	Philippines	Eats peanut weekly	No allergic reaction	0	0.09	0	0.05	0	Yes	Moderate	Yes	Grass pollen allergy
	NA	M	8	Caucasian	Eats peanut weekly	No allergic reaction	0	0.01	0.04	0.01	N/A	No	N/A	No	Cow's milk allergy
	NA	M	6	Caucasian	Eats peanut weekly	No allergic reaction	0	0.01	0.04	0	0	No	N/A	No	Non allergic rhinitis

198

	Peanut allergic status	Sex	Age (yrs)	Ethnicity	Last known peanut exposure	Allergic reaction	Skin prick test (mm)	Peanut specific IgE (kU/L)	Ara h 2	Ara h 1	Ara h 8	Current eczema	Eczema severity	Asthma	Other atopic disease
Flow-cytometry analysis for intracellular cytokine staining	PA	F	5	Caucasian	4 weeks	Angioedema and urticaria	9	1.2	0	1.25	0	Yes	Moderate	Yes	Oral allergy syndrome (OAS), kiwi allergy
	PA	F	15	Caucasian	10 years	Angiodema and hives	8	11.1	N/A	N/A	N/A	Yes	Mild	Yes	OAS, latex allergy, egg, wheat, sesame, mustard, fish allergy
	PA	M	12	Caucasian	9 years	Contact angiodema	15	38.7	11.8	22.3	0.01	No	N/A	Yes	Tree nut and chick pea allergy
	PA	M	14	Caucasian	9 years	Positive oral food challenge	10	2.03	2.14	<0.35	1.1	No	N/A	Yes	Nil else
	PA	F	11	Mixed Black - Caucasian	4 years	Angiodema, hives, difficulty breathing	9	81.2	32.1	42.3	22.8	Yes	Mild	Yes	Tree nut allergy and AR
	PS	M	8	Mixed Black - Caucasian	1-2 times / month	Nil	N/A	7.81	0.18	0.19	1.3	Yes	Moderate	No	Cashew, pistachio and egg allergy
	PS	M	14	Caucasian	Daily	Nil	0	0.38	0	0.01	1.48	Yes	Mild	Yes	Egg, cow's milk and kiwi allergy
	PS	F	6	Black	< 1 week	Nil	0	2.28	0.06	0.18	0.1	Yes	Severe	No	House dust mite allergy
	PS	M	9	South East Asian	<1 week	Nil	0	3.16	0.26	1.79	6.18	N/A	Yes	Yes	Egg, tree-nut+ sesame allergy, tree pollen allergy
	PS	M	10	Caucasian	1 week	Nil	0	3.16	0.26	1.79	6.18	No	N/A	Yes	Egg, tree-nut+ sesame allergy, tree pollen allergy

	Peanut allergic status	Sex	Age (yrs)	Ethnicity	Last peanut exposure	Allergic reaction	SPT (mm)	Peanut sIgE (kU/L)	Ara h 2	Ara h 1	Ara h 8	Current eczema	Eczema severity	Asthma	Atopic disease
Reverse transcriptase quantitative PCR analysis – Peanut allergic	PA	M	11	Other-Caucasian	4 years	Angioedema with Crunchy nut cornflakes	14	>100	>100	>100	76	No	N/A	No	Tree nut allergy allergic rhinitis
	PA	F	6	Caucasian	10 months	Labial / facial angioedema, quiet	8	N/A	0.05	0.78	0	No	N/A	No	Nil
	PA	M	5	Caucasian	1 year	Hives / facial angioedema after peanut butter	10	13.1	9.17	6.45	2.1	No	N/A	No	Tree nut allergy, Oral allergy syndrome, AR,
	PA	M	7	Caucasian	1 year	Hives, facial swelling after cutaneous contact	11	27.1	5.63	0.12	>100	No	N/A	No	Tree nut, sesame, kiwi allergy, OAS, AR
	PA	M	6	Asian	1 year	Facial hives after peanut consumption	10	10.3	1.22	0.9	N/A	Yes	Mild to moderate	No	Cow's milk and egg allergy, AR
	PA	M	7	Black-British	4 years	Angioedema and hives	12	31.8	13.3	38	27	Yes	Moderate	Yes	Fish, tree nut and sesame allergy, AR
	PA	M	3	Caucasian	2 months	DBPCFC: >3 hives after 4.35g peanut protein	5	2.67	2.94	0.03	0.01	No	N/A	No	Nil else
	PA	M	6	Caucasian	1 year	Facial urticaria and vomiting	14	>100	65.5	>100	0.11	Yes	Moderate	Yes	Tree nut, kiwi and egg allergy, AR
	PA	F	10	Other-Caucasian	1 year	Vomiting and generalised hives after chocolate peanuts	12	>100	>100	32	N/A	No	N/A	No	Tree nut and green pea sensitised
	PA	M	8	Asian	5 years	Vomiting, wheeze, difficulty breathing	10	56.5	35	6.6	5.2	No	N/A	No	Tree nut and fish allergy, AR
	PA	M	15	Caucasian	14 years	Cough, vomit, general hives	22	>100	40	68	3.6	No	N/A	No	Tree nut sensitised
	PA	M	6	Caucasian	5 months	General urticaria, vomiting, drowsy post Crunchy nut	12	>100	>100	25.4	0.02	No	N/A	No	Tree nut sensitised

200

	Peanut allergic status	Sex	Age (yrs)	Ethnicity	Last peanut exposure	Allergic reaction	SPT (mm)	Peanut sIgE (kU/L)	Ara h 2	Ara h 1	Ara h 8	Current eczema	Eczema severity	Asthma	Atopic disease
RT-qPCR analysis – Peanut sensitized but tolerant	PS	F	2	Caucasian	<1 month	Negative DBPCFC	4	0.09	0	0	0.01	Yes	Mild	No	Egg allergy
	PS	F	1	Caucasian	<1 month	Negative DBPCFC	4	0.41	0.2	0.01	0.01	Yes	Moderate-severe	No	Egg allergy
	PS	F	6	Asian	<1 month	Negative DBPCFC	2	2.35	0.1	0.1	0.05	No	N/A	No	Egg and fish allergy,
	PS	F	3	SE Asian	<1 month	Negative DBPCFC	3	0.09	0.2	0.01	0.01	Yes	Mild	No	Egg and sesame seed allergy
	PS	M	6m	SE Asian	<1 month	Negative DBPCFC	2	0.74	1.2	0.9	N/A	Yes	Mild to moderate	No	Cow's milk and egg allergy, AR
	PS	M	7	Asian	<1 month	Negative DBPCFC	5	1.86	0.1	0.06	0.06	Yes	Moderate	No	Egg allergy, sensitised to tree nuts
	PS	M	2	SE Asian	<1 month	Negative DBPCFC	5	9.9	0.1	1.4	0.02	No	N/A	No	Cauliflower, fish, shellfish, chickpea, lentil, sesame, egg allergy
	PS	M	3	Caucasian	<1 month	Negative DBPCFC	4	0.83	0.2	0.01	0.01	No	N/A	No	Resolving egg allergy, tree nut allergy
	PS	M	4	Black-Caucasian	2-3 times /week	Nil	4	11.2	0.1	0.47	12.5	Yes	Mild-moderate	Yes	Cow's milk, egg, avocado, banana, sesame allergy, AR
	PS	F	8	Asian	1-2 times/month	Nil	3	3.04	0.3	0.3	N/A	No	N/A	No	Egg, sesame seed allergy
	PS	M	1	Caucasian	<1 month	Negative DBPCFC	6	2.29	0	0.01	0.01	No	N/A	No	Cow's milk allergy
	PS	M	2	Caucasian	<1 month	Negative DBPCFC	2	0.52	0	0.1	0.01	No	N/A	Yes	Tree nut, sesame seed allergy

201

	Peanut allergic status	Sex	Age (yrs)	Ethnicity	Last known peanut exposure	Allergic reaction	Skin prick test (mm)	Peanut sIgE (kU/L)	Ara h 2	Ara h 1	Ara h 8	Current eczema	Eczema severity	Asthma	Atopic disease
RT-qPCR analysis – atopic non-peanut allergic	NA	F	12	Caucasian	1-2 times/month	No allergic reaction	0	0.09	0.02	0.00	0.01	No	N/A	No	Recurrent lip angioedema
	NA	M	12	Caucasian	1-2 times /week	No allergic reaction	0	0.08	0.01	0.08	0.02	No	N/A	No	House dust mite allergy
	NA	M	6	Caucasian	1-2 times/month	No allergic reaction	0	0.01	0.04	0.01	0.01	No	N/A	No	AR, Rhinosinusitis
	NA	M	9	Caucasian	1-2 times/month	No allergic reaction	0	0.18	0.19	0.01	0.01	Yes	Mild	Yes	Resolving egg allergy, kiwi allergy, AR
	NA	M	8	Caucasian	1-2 times /week	No allergic reaction	0	0.01	0.04	0.01	x	No	N/A	No	Cow's milk allergy
	NA	M	8	Philippines	1-2 times /week	No allergic reaction	0	0.09	0	0.05	0	Yes	Moderate	Yes	Grass pollen allergy

202

203 Table EII. Number of cells sorted for each condition in 3 PA and 3 NA individuals and amount and
204 quality of RNA and cDNA obtained from these samples for microarray analysis. (N/A: not available)

Allergy	Culture condition	Activation	Homing	Cells sorted	RNA (ng)	RIN	cDNA ng/µl	cDN A µg	260/ 280	260/ 230
PA1	P+	CD69+	B7	39318	5.8	8.2	225.25	6.8	1.98	1.91
PA1	P+	CD69-	B7	244316	9.0	8.8	253.46	7.6	1.98	1.94
PA1	P+	CD69-	CLA	50193	4.0	8.8	228.95	6.9	1.97	1.87
PA1	P+	CD69+	CLA	12161	2.0	8.0	169.28	5.1	1.96	1.78
PA1	P-	CD69-	B7	38472	5.3	7.9	296.21	8.9	1.97	1.89
PA1	P-	CD69-	CLA	35134	5.5	8.7	239.73	7.2	1.97	1.85
PA2	P+	CD69+	B7	53205	4.5	8.6	255.98	7.7	1.94	1.74
PA2	P+	CD69-	B7	584788	22.1	8.5	265.32	8.0	1.94	1.76
PA2	P+	CD69-	CLA	40685	2.7	9.1	176.84	5.3	1.95	1.78
PA2	P+	CD69+	CLA	12678	0.4	N/A	175.87	5.3	1.96	1.71
PA2	P-	CD69-	B7	44709	0.7	8.1	230.68	6.9	1.93	1.62
PA2	P-	CD69-	CLA	29634	1.7	9.3	186.88	5.6	1.97	1.84
PA3	P+	CD69+	B7	35069	3.9	8.0	185.11	5.6	1.98	1.76
PA3	P+	CD69-	B7	546030	6.0	N/A	265.82	8.0	1.95	1.81
PA3	P+	CD69-	CLA	478926	17.8	9.0	192.68	5.8	1.97	1.88
PA3	P+	CD69+	CLA	55106	3.3	8.3	138.93	4.2	1.96	1.73
PA3	P-	CD69-	B7	58439	3.4	8.3	131.97	4.0	1.97	1.83
PA3	P-	CD69-	CLA	148262	10.1	8.2	151.99	4.6	1.98	1.84
NA1	P+	CD69+	B7	40271	3.9	8.4	245.45	7.4	1.95	1.76
NA1	P+	CD69-	B7	669075	3.6	7.9	258.61	7.8	1.95	1.70
NA1	P+	CD69-	CLA	228356	4.1	8.4	244.64	7.3	1.94	1.69
NA1	P+	CD69+	CLA	17997	1.3	N/A	175.56	5.3	1.93	1.55
NA1	P-	CD69-	B7	84218	2.5	8.2	215.43	6.5	1.96	1.71
NA1	P-	CD69-	CLA	76554	4.6	8.3	240.61	7.2	1.96	1.77
NA2	P+	CD69+	B7	34785	8.3	7.2	189.08	5.7	2.00	1.99
NA2	P+	CD69-	B7	347340	15.9	8.8	265.95	8.0	1.99	1.96
NA2	P+	CD69-	CLA	90357	8.1	7.3	264.92	7.9	1.98	1.94
NA2	P+	CD69+	CLA	8505	1.2	7.2	110.82	3.3	1.96	1.81
NA2	P-	CD69-	B7	94715	8.0	8.6	219.77	6.6	1.97	1.90
NA2	P-	CD69-	CLA	56690	6.7	6.3	242.87	7.3	1.98	1.91
NA3	P+	CD69+	B7	759119	7.0	8.5	246.9	7.4	1.91	1.62
NA3	P+	CD69-	B7	40447	1.9	7.5	248.84	7.5	1.93	1.62
NA3	P+	CD69-	CLA	124397	5.0	8.4	291.68	8.8	1.95	1.74
NA3	P+	CD69+	CLA	9854	1.4	N/A	161.63	4.8	1.91	1.51
NA3	P-	CD69-	B7	73489	3.6	8.9	290.01	8.7	1.95	1.78
NA3	P-	CD69-	CLA	45974	2.6	8.7	279.49	8.4	1.95	1.80

205

206

Online repository legends:

FIG. E1. FACS sorting experiments for 3 PA and 3 NA donors illustrating the isolation of memory Th cells (expressing CD4 and CD45RO) homing to the skin (CLA+) or the gut (B7+) and either peanut-activated (CD69+) or non-activated controls (CD69-).

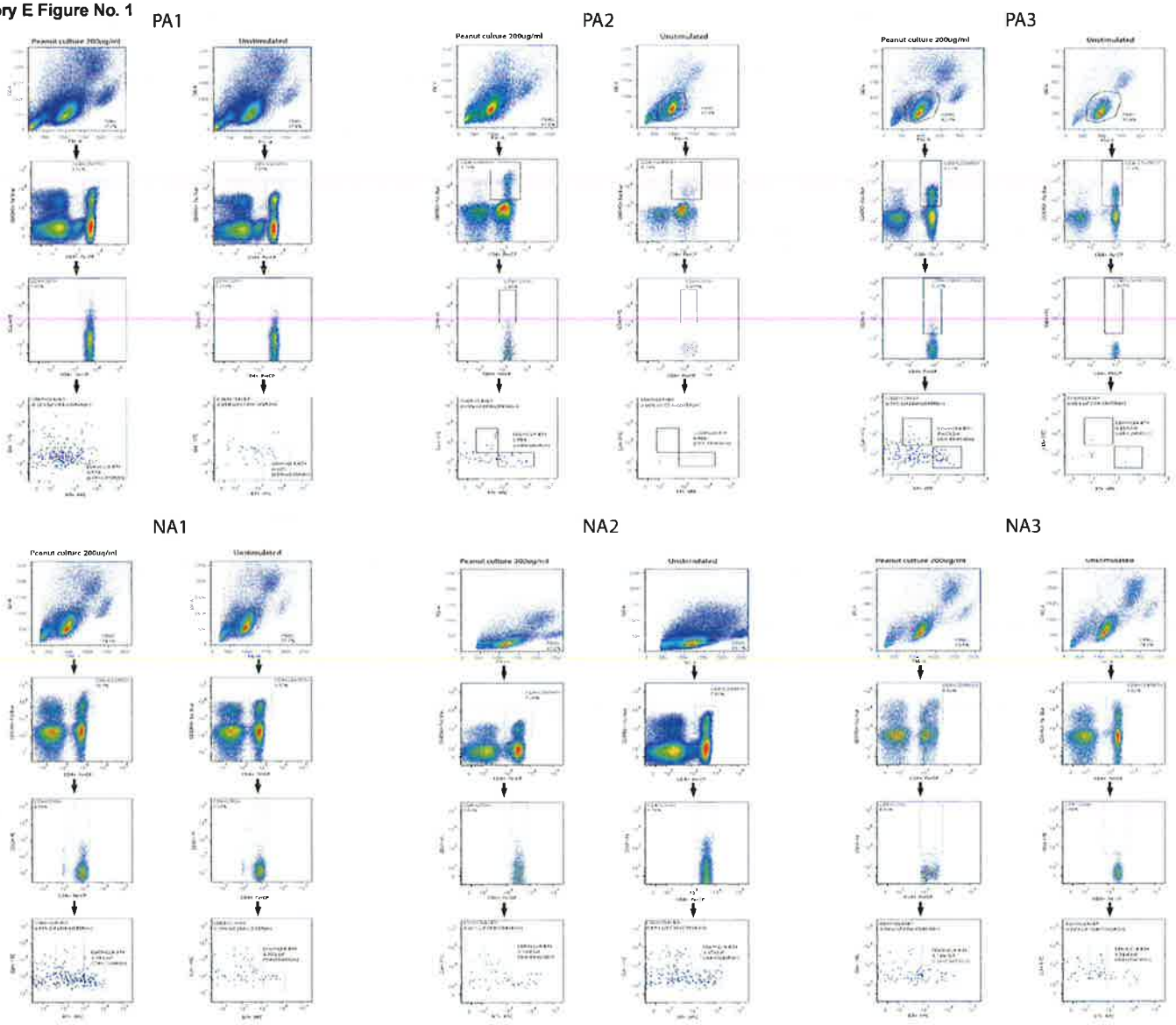
FIG E2. Data normalization and quality checks for microarray hybridization. Box plots of array hybridization signal before and after RMA normalization were shown in FIG Ai and Aii, respectively. All arrays had a uniform pattern of hybridization control signal (FIG B), which was at the expected intensity in congruent with their staggered concentration of CreX>BioD>BioC>BioB. Red: CreX, blue: BioD, pink: BioC, green: BioB.

FIG. E3. Principal Component Analysis of Gene expression in the Affymetrix GeneChip® Human Gene1.0 ST microarray led to clustering of the CLA+ skin homing Th cells (green centrinoid) distinguishing them from B7 gut homing Th cells (yellow centrinoid).

FIG. E4. Gene expression in peanut stimulated (Stim) versus unstimulated (Unstim) culture conditions in activated (CD69+) and non-activated (CD69-) skin (CLA+) and gut (B7+) homing memory Th cells in PA versus NA children.

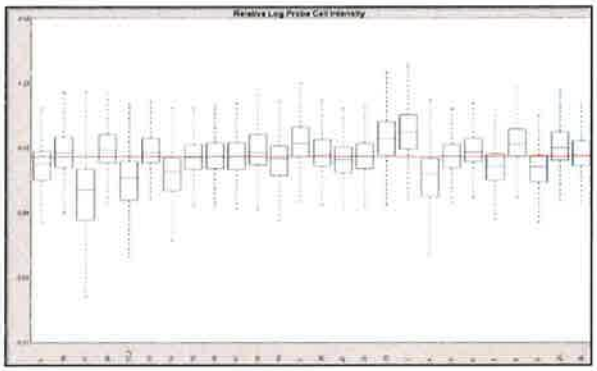
FIG. E5. Random Forest algorithm for identifying classifier genes which differentiate between peanut-activated memory Th cells from PA and NA donors.

Repository E Figure No. 1

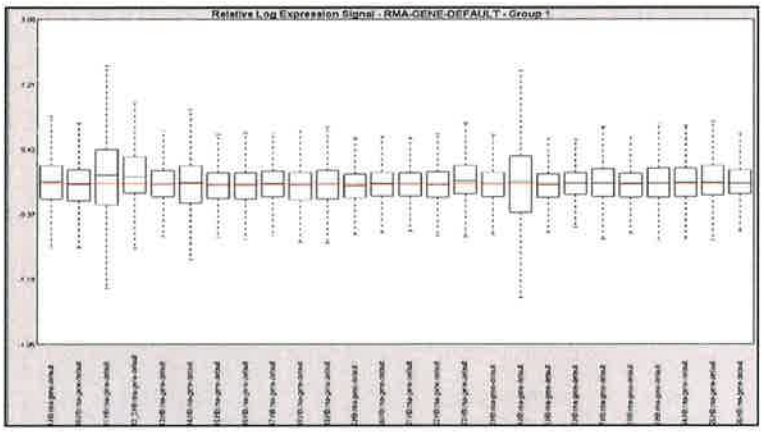


Repository E Figure No. 2

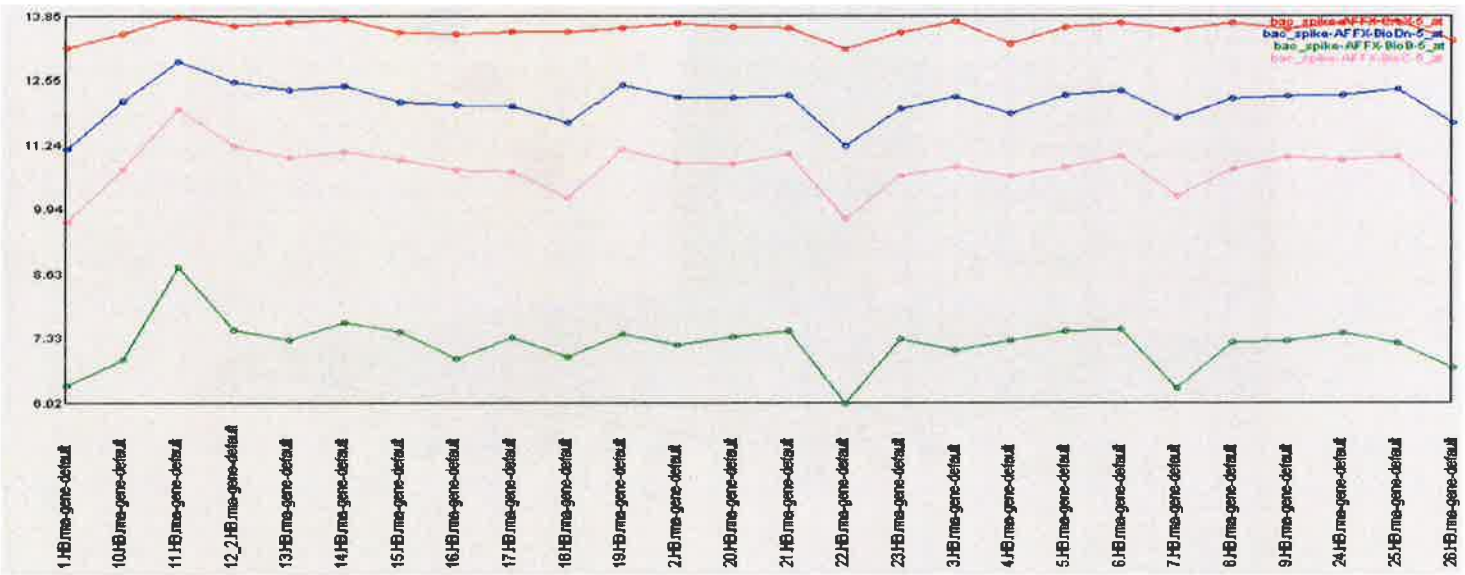
Ai. Before normalisation:



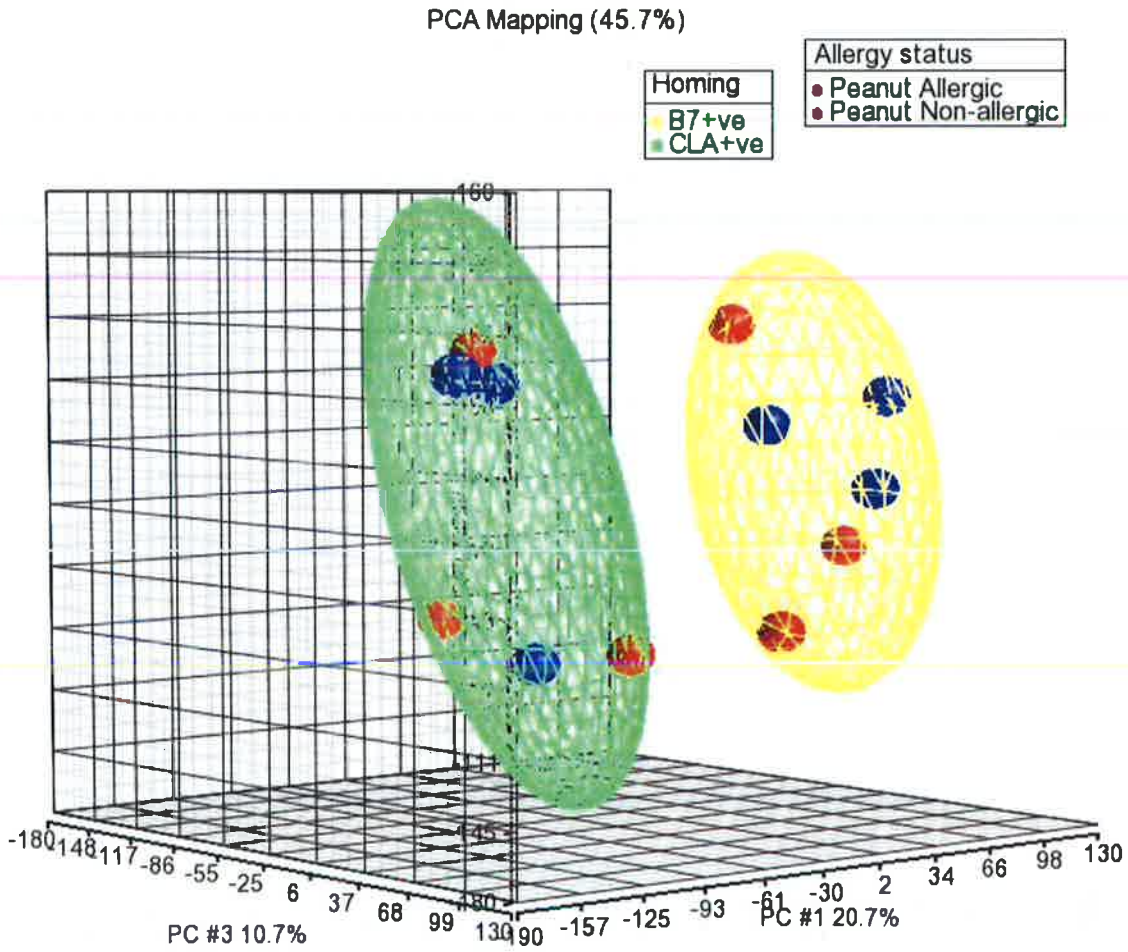
Aii. After RMA normalisation:



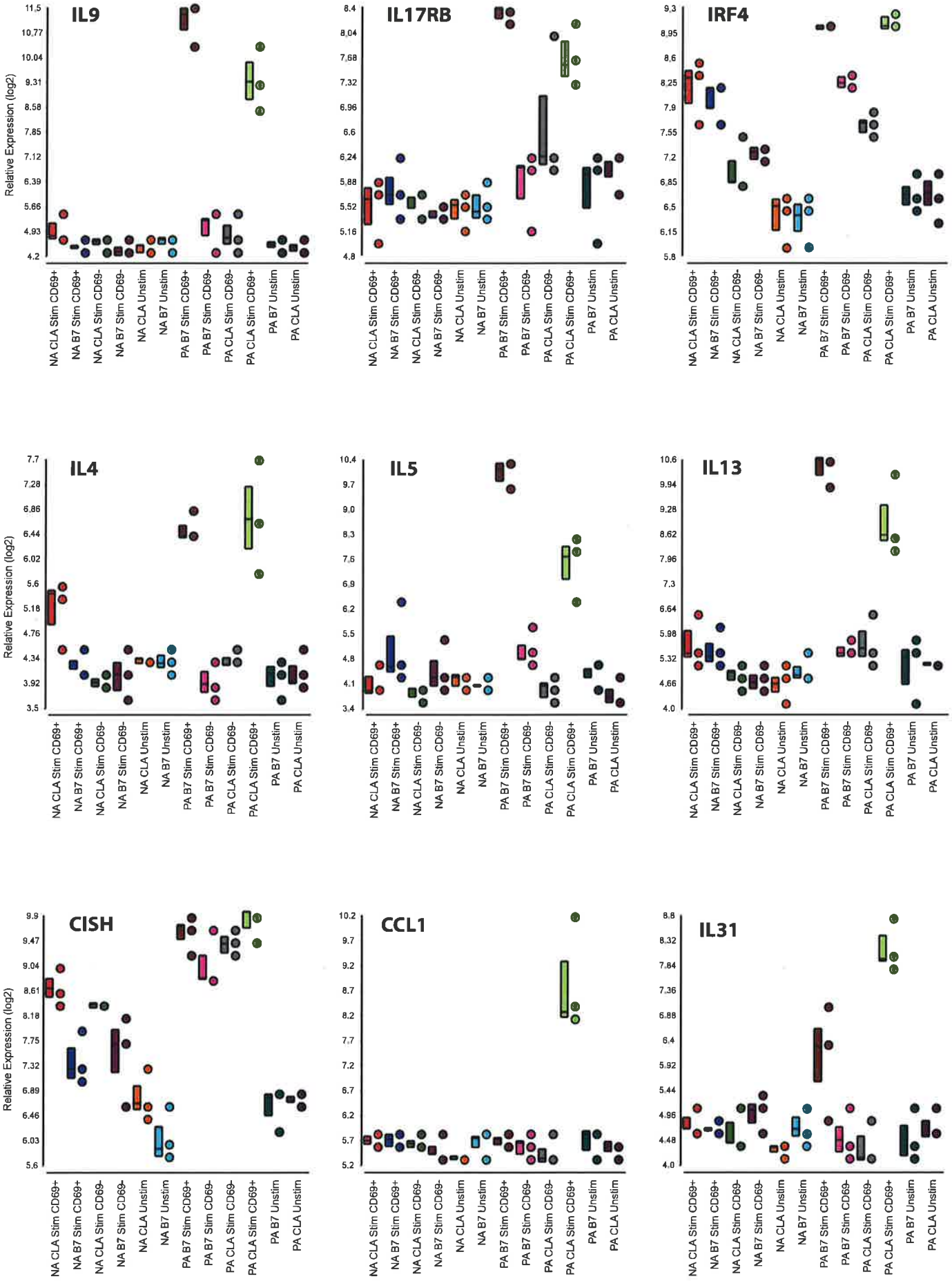
B. Hybridisation control



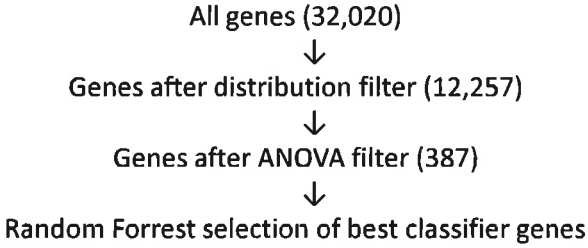
Repository E Figure No. 3



Repository E Figure No. 4



Repository E Figure No. 5



Genes after distribution filter: there are 3 possible 2-gene combinations which all classify with 100% accuracy:

1) 750 trees:
   IL9
   HUWE1 (HECT, UBA and WWE domain containing 1)

2) 500 trees:
   IL9
   8069610 (hypothetical LOC100271832; RNA, Ro-associated Y5 pseudogene 10; RNA, Ro-associated Y1; RNA, Ro-associated Y4 pseudogene 7; RNA, Ro-associated Y4 pseudogene 19; RNA, Ro-associated Y3; hypothetical LOC100132111; RNA, Ro-associated Y4)

3) 250 trees:
   IL9
   HSPA5 (LOC400750 hypothetical gene supported by AF216292; NM_005347; heat shock 70kDa protein 5, glucose-regulated protein, 78kDa)